Using Mask R-CNN to perform crop field identification

A Report Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the
Degree of Master of Professional Studies in
Agriculture and Life Sciences
with a specialization in Geospatial Applications

by
Yuchen Xie
August 2020

ABSTRACT

In recent years, the usage of machine learning algorithms has been observed in multiple fields of work, ranging from automated driving capabilities and facial recognition technologies. One of the most recently developed algorithm is called Mask R-CNN, which is an extension of the Faster R-CNN architecture that serves to quickly perform transfer learning on various training sample sets based on a larger pool of existing data.

In this project, I will be using Mask R-CNN to better understand land features, namely crop field identification. This allows a much quicker way of extracting existing crop field locations via satellite imagery as compared to manual identification. A total of 86 low-resolution satellite images were collected for this project, using a 4:1 training and validation ratio for accuracy determination.

The entire machine learning process took one week, which produced a result of around 78% accuracy.

Biological Sketch

Yuchen Xie is from Singapore, a small city-state with very few opportunities to enjoy the presence of natural sceneries. As a result, Yuchen always had an interest in learning about the various geological formations on Earth, resulting in his degree in Geography at the National University of Singapore (NUS).

In NUS, Yuchen developed a passion for geospatial studies after taking a GIS class, learning deeply about various remote sensing techniques and big data management. He has constantly sought ways to better study the huge amounts of remotely sensed data, eventually tapping upon the idea of utilizing machine learning algorithms to do so via automated processes that would seek to save plenty of time.

After much discussion with Dr. Ying Sun, he was finally able to be enrolled into Cornell's MPS degree with a specialization in Geospatial Applications, allowing him to take more advanced lessons in this field of work. These classes were critical in refining his skills in both GIS and remote sensing techniques immensely.

Acknowledgments

I would like to thank my advisor, Dr. Ying Sun, for her patience and understanding in guiding me through my MPS degree in Cornell. She was not just an incredible advisor when commenting on aspects of my project that requires improvements, but also helping me make better decisions in the classes that I should take that would benefit me in achieving my goals.

I would also like to thank everyone in the Sun Lab, who had been so helpful especially in the administrative aspects of my degree. I would not have been able to collect the data required for this project without their help.

Finally, I would like to say a huge thank you to the staff working at the Cornell Institute for Social and Economic Research (CISER). My academic journey in Cornell was unfortunately hit by a global pandemic which closed off the entire campus, denying me access to much of the school's resources. This was a huge blow for my project since the nature of it requires me to gain access to computers with relatively powerful computing capabilities, in which my laptop would not have been able to handle. The staff working at CISER was kind enough to grant me access to

one of their online cloud servers completely free of charge, allowing me to have the possibility of continuing my project. Despite a major server reboot that wiped out much of my project, one of the staff members, Jacob Grippin, tried his best to attempt to recover my work, in which I will forever be grateful for.

Table of Contents

1. Introduction to machine learning methods

The prevalence of machine learning algorithms developed since the past decade has been incorporated in multiple fields that enhances the efficiency for work that used to be done manually via human effort. With the ability to automate manual work and its capability to conduct constant recalibrations to improve its accuracy and allow a reduction in committed errors, researchers and technical specialists have constantly been trying to perfect their systems with better computational algorithms to improve on both of those processes. One of the most recent algorithms developed is Mask R-CNN, and advanced variation of Faster R-CNN, which is most known for being applied in facial recognition technologies. For this project, I will be attempting to apply Mask R-CNN on land use classification instead.

Faster R-CNN (Ren, He, Girshick & Sun, 2017) is a bounding box detector that boxes out and labels recognizable objects in an input image. Its initial implementation uses a VGG-16 feature extractor which is a 16 layered convolutional neural network. The features extracted are passed into two subsequent modules along the pipeline. The features are first passed into a Region Proposal Network (RPN), which attempts to use fit a set of pre-trained anchor boxes to box out possible interesting regions in the image without labelling them. Every region of the feature image boxed out by the RPN is then passed into an image classifier, which identifies the object within the box using the feature and classifies them. Combined, the output image shows the boxes identified by the RPN with classes labelled by the classifier

Mask R-CNN (He, Gkioxari, Dollar & Girshick, 2020) is built on top of the Faster R-CNN infrastructure. The goal of Mask R-CNN is to perform binary segmentation - labeling all pixel belonging to each instance of an identified object. After passing the features through the RPN and classifier, Mask R-CNN takes each identified object and refines the size of the bounding box using knowledge of the class of object contained in the box, before performing binary segmentation on pixels within the refined bounding box. In the output of Mask R-CNN, each object identified is then tagged to a set of labelled pixels instead of a bounding box.

2. Usage of machine learning algorithms in crop classification

There have been various papers published that applied machine learning algorithms in land use classification in the past. For example, recurrent neural network (RNN) have been applied onto temporal patterns of crops across a time series to obtain crop field identification via a longitudinal study (Sun, Di & Fang, 2018). Combined with existing feature extraction tools and compared with ground truth data via CDL

(Cropland Data Layer) datasets (https://nassgeodata.gmu.edu/CropScape/), an accuracy result of 88% was achieved. However, I was unable to apply this method into my project due to the vast amount of time an additional temporal dimension would require to run the algorithm properly. Instead, I will be comparing my results with the paper published by Wang et al. (Wang, Azzari & Lobell, 2019), which applies random forest transfer and unsupervised training for crop-type mapping. Overall, this paper was able to achieve over 85% accuracy on low crop diversity areas, but almost no better than random in high crop diversity areas. I would be evaluating my project based on comparisons made with these results to get a sense of Mask R-CNN's applicability and usefulness when conducting land use classification.

3. Data used

For this study, I will be using satellite imagery collected from Sentinel 2A and 2B as training data. Despite its relatively low spatial resolution at only 10-60m, it is still advantageous to use imagery from Sentinel over commercial satellite data due to its low cost and quantity available. Since most machine learning algorithms require huge amounts of images to serve as training data, it is not feasible to purchase satellite images from commercial companies.

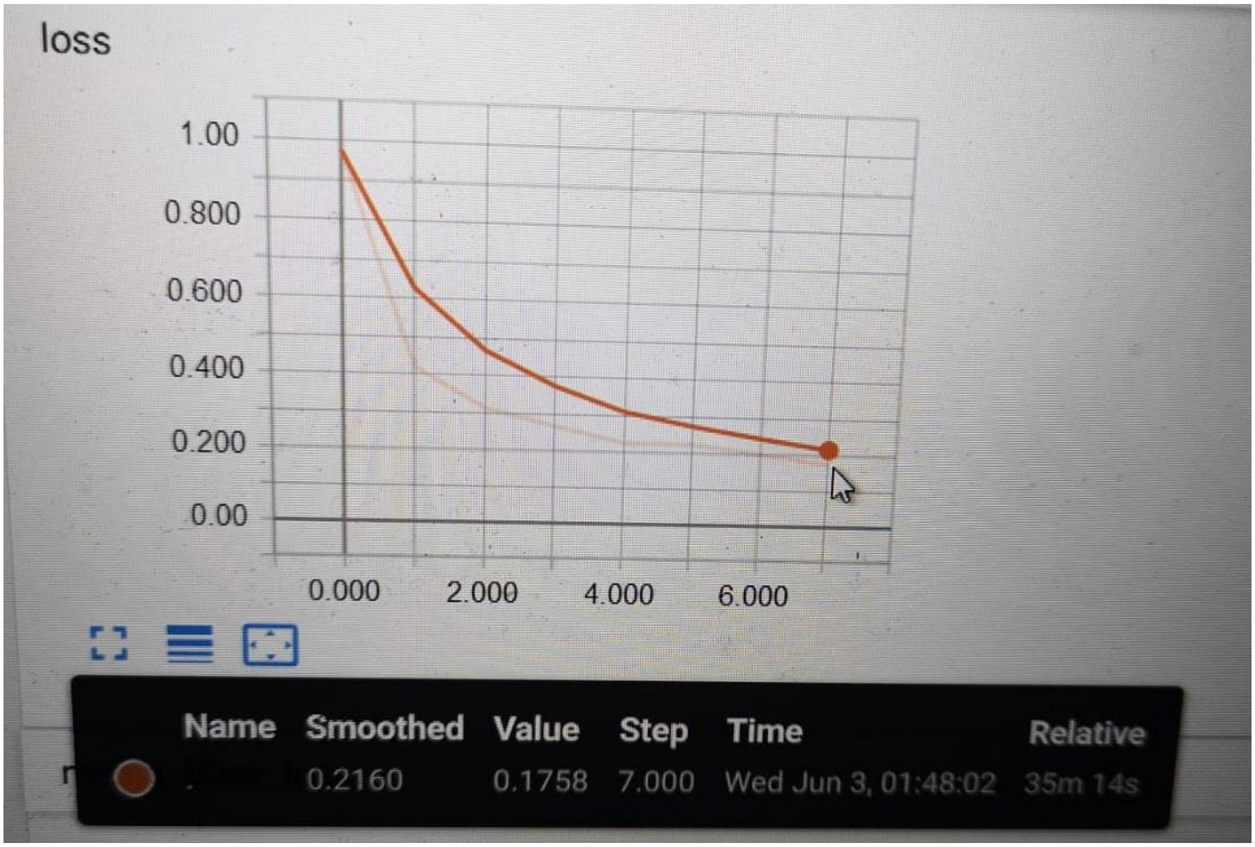
As for the nature of the satellite imagery used, there were a few criteria implemented to ensure consistent crop growth and image quality. Firstly, satellite imagery purely focusing on New York state was used. Secondly, the temporal period of 2010-2019 was selected, with images only taken in September used for this project. This is because the September period tends to be the peak crop season for two major crop types, Corn and Soybean, thus ensuring more distinctive features are present within crop fields for feature extraction during the machine learning process. Finally, a cloud cover below 10% was ensured for the best image quality.

This amounted to a total of 86 satellite images used for this project. Since the sample size is still relatively small, I will be using a 4:1 ratio to segregate training and validation data. This amounts to 64 images that will serve as training samples and 22 images used as validation data to test the accuracy of the trained result. For the images within the training set, manual visual identification along with ground truth information collected from the CDL was applied so as to identify the major crop fields.

4. Segregating crop fields

The attempt at land use classification using Mask R-CNN would be tested by attempting to differentiate major landforms that encompass greater land areas. As mentioned above, numerous projects have been conducted in the past to establish accurate land use land cover maps with different areas of interest. For this project, the ability to identify crop fields would be a crucial component in testing the capabilities of Mask R-CNN in dissecting specific objects from various satellite images.

In order to ensure optimal training images were selected to extract crop fields, the following criteria were applied during the selection process. Firstly, the total area coverage for buildings within one satellite image cannot exceed 80%. This is to ensure that the algorithm is able to pick up existing feature differences between crop fields and non-crop fields properly, which would not have been possible if an overwhelming huge proportion of the image is covered by crop fields in the first place. Secondly, satellite imagery that contains large numbers of urban landscape features such as buildings, while also having almost no crop field land covers will have to be avoided as well. This is because the amount of urban landscape tends to be that of an inverse proportion to the area covered by crop fields, which may result in features such as buildings being used as an important feature that determines crop field identification even though they are separate land features in the first place. As a result, six images were rejected from being part of the training dataset.



	Name	Smoothed	Value	Step	Time	Relative
●	[illegible]	0.2160	0.1758	7.000	Wed Jun 3, 01:48:02	35m 14s

Figure 1. 0.22% loss based on 7 epoch cycles.

After a week of running Mask R-CNN over the training images along with cross-validation with the rest of the images over 7 epoch cycles, the result amounted to an error rate of around 0.22 (Fig 1). This means that around 78% of the time, the crop fields were correctly identified and recognized based on the Mask R-CNN algorithm.

5. Result

Comparing with the Wang et al. paper, I was unable to achieve better than 85% accuracy which was the initial benchmark I set for this project. The primary reason could be attributed to the lack of epoch cycles and amount of training data used. Given the luxury of more time and better computing hardware capabilities, the result would very likely be higher than 78% accuracy. Another reason could be the lack of land coverage, since the images I selected for this project was only based in New York, while Wang et al. collected satellite images from multiple states. Finally, Wang at al. had a better crop diversity, since other crops such as wheat and alfalfa were included in their study. This provides better attribute distinction for crop field identification, which is significantly better than this project which was only based on two crop types.

6. Conclusion

The purpose of this project is to identify whether Mask R-CNN can be effectively utilized for land cover identification. In this sense, the results achieved despite the lack of time and data quantity is satisfactory enough to label it as being relatively successful. Even though manual identification may still be more accurate, the ability for Mask R-CNN to be able to conduct fully automated processes allow researchers and urban planners to divert more time and resources into other aspects of landscape management instead of manual time spent in land feature identification. It is hoped that in the future, better machine learning algorithms can be developed to refine upon current feature extracting technologies, improving vastly on both accuracy and time spent in all fields of work.

References

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2020). Mask R-CNN. IEEE Transactions On Pattern Analysis And Machine Intelligence, 42(2), 386-397. doi: 10.1109/tpami.2018.2844175

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions On Pattern Analysis And Machine Intelligence, 39(6), 1137-1149. doi: 10.1109/tpami.2016.2577031

Sun, Z., Di, L., & Fang, H. (2018). Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. International Journal Of Remote Sensing, 40(2), 593-614. doi: 10.1080/01431161.2018.1516313

Wang, S., Azzari, G., & Lobell, D. (2019). Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. Remote Sensing Of Environment, 222, 303-317. doi: 10.1016/j.rse.2018.12.026